Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models

Jinho Baik* and **Jack W. Silverstein**†

July 27, 2004

Abstract

We consider a spiked population model, proposed by Johnstone, whose population eigenvalues are all unit except for a few fixed eigenvalues. The question is to determine how the sample eigenvalues depend on the non-unit population ones when both sample size and population size become large. This paper completely determines the almost sure limits for a general class of samples.

1 Introduction

The sample covariance matrix is fundamental to multivariate statistics. When the population size is not large and for a sufficient number of samples, the sample covariance matrix is a good approximate of the population covariance matrix . However when the population size is large and comparable with the sample size, as is in many contemporary data, it is known that the sample covariance matrix is no longer a good approximation to the covariance matrix. The Marchenko-Pastur theorem [15] states that with $n =$ the sample size, $p =$ the population size, as $n = n(p) \to \infty$ such that $\frac{p}{n} \to c$, the eigenvalues $s_j^{(p)}$, $j = 1, \ldots, p$, of the sample covariance matrix of normalized i.i.d. Gaussian samples satisfy for any real x

$$\frac{1}{p}\#\{s_j^{(p)} : s_j^{(p)} < x\} \to F(x) \tag{1.1}$$

almost surely where

$$F'(x) = \frac{1}{2\pi x c}\sqrt{(b-x)(x-a)}, \qquad a < x < b, \tag{1.2}$$

and $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$ when $0 < c \le 1$. When $c > 1$, there is an additional Dirac measure at $x = 0$ of mass $1 - \frac{1}{c}$. Moreover, there are no stray eigenvalues in the sense that the top and bottom eigenvalues converge to the edges of the support of F [8]:

$$s_1^{(p)} \to (1+\sqrt{c})^2 \tag{1.3}$$

*Department of Mathematics, University of Michigan, Ann Arbor, MI, 48109, USA, baik@umich.edu

†Department of Mathematics, North Carolina State University, Raleigh, NC, 27695, USA, jack@math.ncsu.edu

AMS 1991 *subject classifications.* Primary 15A52, 60F15; secondary 62H99.

Key words and phrases. Eigenvalues, sample covariance matrices, spiked population models, almost sure limits, non-null case.

almost surely and [20]

$$s^{(p)}_{\min\{p,n\}} \to (1-\sqrt{c})^2 \tag{1.4}$$

almost surely ($s^{(p)}_{n+1} = \ldots s^{(p)}_p = 0$ when $n < p$). One can extract from this some information of the population covariance matrix even though the sample covariance matrix is not a good approximate. For example, if there are non-zero eigenvalues of the sample covariance matrix well separated from the rest of the eigenvalues, one finds, assuming the Gaussian entries, that the samples are not i.i.d..

There are indeed many cases in which a few eigenvalues of the sample covariance matrix are separated from the rest of the eigenvalues, the latter being packed together as in the support of the Marchenko-Pastur function (1.2). The examples include speech recognition [6, 12] , mathematical finance [18], [13], [14], wireless communication [25], physics of mixture [19], and data analysis and statistical learning [10].

As a possible explanation for such features, Johnstone [12] proposed the 'spiked population model' where all but finitely many eigenvalues of the population covariance matrix are the same, say equal to 1. The question is how the eigenvalues of the sample covariance matrix would depend on the non-unit population eigenvalues as $p, n \to \infty$. It is known [15, 22] that the Marchenko-Pastur result (1.1) still holds for the spiked model. But (1.3) and (1.4) are not guaranteed and some of the eigenvalues are not necessarily in the support of (1.2).

For example, consider the case when the population covariance matrix has one non-unit eigenvalue, denoted by σ_1. If σ_1 is close to 1, one would expect that as the dimension p becomes large, the population covariance matrix would be close to a large identity matrix, and hence σ_1 would have little effect on the eigenvalues of the sample covariance matrix. On the other hand, if σ_1 is much bigger than 1, then even if p becomes large, σ_1 might still pull up the eigenvalues of the sample covariance matrix. How big should σ_1 be in order to have any effect, how many eigenvalues of the sample covariance matrix would be pulled up and exactly where would the pulled-up eigenvalues be ? We will see in the results below that the answers are $\sigma_1 > 1+\sqrt{c}$ (where $\frac{p}{n} \to c$), one eigenvalue at most, and $\sigma_1 + \frac{c\sigma_1}{\sigma_1 - 1}$, respectively.

For *complex Gauassian* samples, the papers [17, 5] study the *largest* eigenvalue of the sample covariance matrix. The authors determine the transition behavior and the limiting distributions are also obtained. The purpose of this paper is a complete study of the spiked model for *both real and complex samples which are not necessarily Gaussian.* We obtain almost sure limit results. A general study of 'non-null' covariance matrices was done in [2, 3]. We will show in this paper how to extract the desired results from the work of [3]. While this paper was being prepared, the authors learned that Debashis Paul [16] was also studying the spiked model independently at the same time, which has some overlap with this work. Paul considers the real Gaussian samples for $c < 1$, and obtains the almost sure limits as in (1.10) and (1.11) below for large sample eigenvalues. Moreover, when all non-unit population eigenvalues are simple, the limiting distribution is found to be Gaussian (see Subsection 1.3 below for more detail). On the other hand, our paper (i) is concerned with more general samples, not necessarily Gaussian, (ii) includes all choices of c and (iii) studies both large and small sample eigenvalues. We remark that a complete study of limiting distributions is still an open question.

1.1 Model

Let T_p be a fixed $p \times p$ non-negative definite Hermitian matrix. Let Z_{ij}, $i, j = 1, 2, \dots$, be independent and identically distributed complex valued random variables satisfying

$$\mathbb{E}(Z_{11}) = 0, \qquad \mathbb{E}(|Z_{11}|^2) = 1, \quad \text{and} \quad \mathbb{E}(|Z_{11}|^4) < \infty, \tag{1.5}$$

and set $Z_p = (Z_{ij})$, $1 \le i \le p$, $1 \le j \le n$. We take the sampled vectors to be the columns of $T_p^{1/2} Z_p$, where $T_p^{1/2} Z_p$ is an Hermitian square root of T_p. Hence T_p is the population covariance matrix. Of course, not all random vectors are realized as such, but this model is still very general. When Z_{ij} are i.i.d (real or complex) Gaussian, the model becomes the Gaussian samples with population covariance matrix T_p. Outside the Gaussian case we see that these vectors cover a broad range of random vectors, completely real or complex, with arbitrary population covariance matrix.

Let

$$B_p := \frac{1}{n} T_p^{1/2} Z_p Z_p' T_p^{1/2} \tag{1.6}$$

be the sample covariance matrix, where Z_p' denotes conjugate transpose. Denote the eigenvalues of B_p by $s_1^{(p)}, \dots, s_p^{(p)}$: for some unitary matrix U_B,

$$U_B B_p U_B^{-1} = \begin{pmatrix} s_1^{(p)} & & & \\ & s_2^{(p)} & & \\ & & \ddots & \\ & & & s_p^{(p)} \end{pmatrix} = diag(s_1^{(p)}, s_2^{(p)}, \dots, s_p^{(p)}). \tag{1.7}$$

For definiteness, we order the eigenvalues as $s_1^{(p)} \ge s_2^{(p)} \ge \cdots \ge s_p^{(p)} \ge 0$.

Let $\alpha_1 > \cdots > \alpha_M > 0$ be fixed real numbers for some fixed $M \ge 0$, which is independent of p and n. Let $k_1, \dots, k_M$ be fixed non-negative integers and set $r = k_1 + \cdots + k_M$, which are also independent of p and n. We assume that all the eigenvalues of T_p are 1 except for, say, the first r eigenvalues. This is the 'spiked population model' proposed in [12]. Let the first r eigenvalues be equal to $\alpha_1, \dots, \alpha_M$ with multiplicity $k_1, \dots, k_M$, respectively: for some unitary matrix U_T,

$$U_T T_p U_T^{-1} = diag(\underbrace{\alpha_1, \dots, \alpha_1}_{k_1}, \underbrace{\alpha_2, \dots, \alpha_2}_{k_2}, \dots, \underbrace{\alpha_M, \dots, \alpha_M}_{k_M}, \underbrace{1, \dots, 1}_{p-r}). \tag{1.8}$$

We set $k_0 = 0$.

1.2 Results

Theorem 1.1 (case $c < 1$). *Assume that $n = n(p)$ and $p \to \infty$ such that*

$$\frac{p}{n} \to c \tag{1.9}$$

for a constant $0 < c < 1$. Let M_0 be the number of j's such that $\alpha_j > 1 + \sqrt{c}$, and let $M - M_1$ be the number of j's such that $\alpha_j < 1 - \sqrt{c}$. Then the following holds.

- *For each $1 \leq j \leq M_0$,*

$$s^{(p)}_{k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \leq i \leq k_j. \tag{1.10}$$

almost surely.

- $$s^{(p)}_{k_1+\cdots+k_{M_0}+1} \to (1+\sqrt{c})^2 \tag{1.11}$$

almost surely.

- $$s^{(p)}_{p-r+k_1+\cdots+k_{M_1}} \to (1-\sqrt{c})^2 \tag{1.12}$$

almost surely (recall $r = k_1 + \cdots + k_M$).

- *For each $M_1 + 1 \leq j \leq M$,*

$$s^{(p)}_{p-r+k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \leq i \leq k_j \tag{1.13}$$

almost surely.

Therefore, when $c < 1$, in order for a population eigenvalue to contribute a non-trivial effect to the eigenvalues of the sample covariance matrix, it should sufficiently big (larger than $1+\sqrt{c}$) or sufficiently small (less than $1-\sqrt{c}$). As an example, when $r = 1$, by denoting the only non-unit eigenvalue by σ_1, the largest sample eigenvalue $s^{(p)}_1$ satisfies

$$s^{(p)}_1 \to \begin{cases} (1+\sqrt{c})^2, & \sigma_1 \leq 1+\sqrt{c} \\ \sigma_1 + \frac{c\sigma_1}{\sigma_1 - 1}, & \sigma_1 > 1+\sqrt{c} \end{cases} \tag{1.14}$$

almost surely. When $r = 2$, by denoting the two non-unit eigenvalues by σ_1, σ_2, the largest sample eigenvalue $s^{(p)}_1$ satisfies

$$s^{(p)}_1 \to \begin{cases} (1+\sqrt{c})^2, & \max\{\sigma_1,\sigma_2\} \leq 1+\sqrt{c} \\ \max\{\sigma_1,\sigma_2\} + \frac{c\max\{\sigma_1,\sigma_2\}}{\max\{\sigma_1,\sigma_2\}-1}, & \max\{\sigma_1,\sigma_2\} > 1+\sqrt{c} \end{cases} \tag{1.15}$$

almost surely.

The results (1.10) and (1.11) are also independently obtained in [16] under the assumption that the samples are Gaussian.

Theorem 1.2 (case $c > 1$). *Assume that $n = n(p)$ and $p \to \infty$ such that*

$$\frac{p}{n} \to c \tag{1.16}$$

for a constant $c > 1$. Let M_0 be the number of j's such that $\alpha_j > 1 + \sqrt{c}$. Then the following holds.

- *For each $1 \leq j \leq M_0$,*

$$s^{(p)}_{k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \leq i \leq k_j. \tag{1.17}$$

almost surely.

- $$s^{(p)}_{k_1+\cdots+k_{M_0}+1} \to (1+\sqrt{c})^2 \tag{1.18}$$

  almost surely.

- $$s^{(p)}_n \to (1-\sqrt{c})^2 \tag{1.19}$$

  almost surely.

- *For all* p,
  $$s^{(p)}_{n+1} = \cdots = s^{(p)}_p = 0. \tag{1.20}$$

Thus, unlike the case of $c < 1$, small eigenvalues of T_p do not affect the eigenvalues of B_p when $c > 1$.

Theorem 1.3 (case $c = 1$). *Assume that $n = n(p)$ and $p \to \infty$ such that*

$$\frac{p}{n} \to 1. \tag{1.21}$$

Let M_0 be the number of j's such that $\alpha_j > 2$. Then the following holds.

- *For each* $1 \le j \le M_0$,
  $$s^{(p)}_{k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{\alpha_j}{\alpha_j - 1}, \qquad 1 \le i \le k_j. \tag{1.22}$$

  almost surely.

- $$s^{(p)}_{k_1+\cdots+k_{M_0}+1} \to 4 \tag{1.23}$$

  almost surely.

- $$s^{(p)}_{\min\{n,p\}} \to 0 \tag{1.24}$$

  almost surely.

1.3 Discussion

As mentioned earlier, the limiting density of the eigenvalues of spiked population models is given by the Marchenko-Pastur theorem (1.2) as in the identity population matrix case, and for the top eigenvalue $s^{(p)}_1$ in the *complex Gaussian* case, the results Theorem 1.1 and Theorem 1.3 were first obtained in [17, 5]. The paper [5] (see section 6) contains an interesting heuristic argument for the critical value $1+\sqrt{c}$ and the value

$$\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \tag{1.25}$$

for $1 \le j \le M_0$ above: they come from a competition between a 1-dimensional last passage time and a 2-dimensional last passage time. It would be interesting to have such a heuristic reasoning for the general case.

When T_p is the identity matrix (the 'null case'), under the Gaussian assumption, the limiting distribution for the largest eigenvalue is obtained for the complex case in [7, 11] and for the real case in [12]. [24] shows the Gaussian assumption is not necessary when $c = 1$. The limiting distributions are the Tracy-Widom distributions [26, 27] in the random matrix theory in mathematical physics. For the spiked model with complex Gaussian samples when $c \leq 1$, the limiting distributions of the largest eigenvalue are obtained in [17, 5]. The paper [5] determines the limiting distribution of $s_1^{(p)}$ for complete choices of the largest population eigenvalue α_1 and its multiplicity k_1: the distribution is (i) the Tracy-Widom distribution when $\alpha_1 < 1 + \sqrt{c}$, (ii) certain generalizations of the Tracy-Widom distribution (see also [4]) when $\alpha_1 = 1 + \sqrt{c}$, and (iii) the Gaussian distribution ($k_1 = 1$) and its generalization ($k_1 \geq 2$, the Gaussian unitary ensemble) when $\alpha_1 > 1 + \sqrt{c}$. For real Gaussian samples [16] showed that when $c < 1$, $M_0 \geq 1$ and $k_1 = \cdots = k_{M_0} = 1$, the limiting distribution of $s_j^{(p)}$, $1 \leq j \leq M_0$, is Gaussian. It is an interesting open question to determine the limiting distribution for the general case of real samples. See section 1.3 of [5] for a conjecture for the scaling.

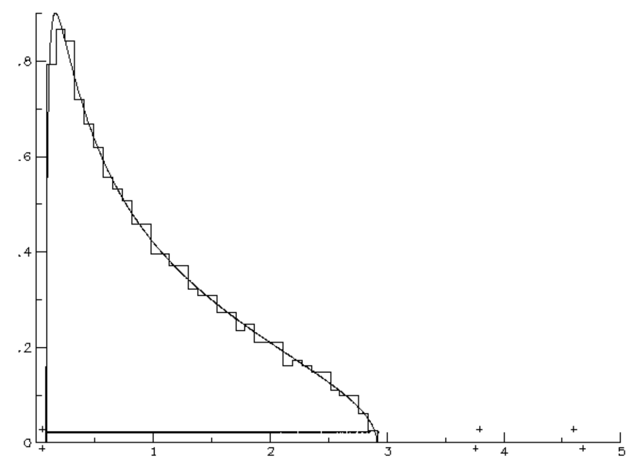

Figure 1: Gaussian samples when $p = 1000, n = 2000$

We include several plots for the case when $c = 0.5$ and there are three non-unit population eigenvalues given by 0.1, 3 and 4 (of multiplicity 1 each). In this case, the critical values of the eigenvalues are $1 + \sqrt{c} \simeq 1.70711$ and $1 - \sqrt{c} \simeq 0.29289$. Hence theoretically we expect that three sample covariance eigenvalues of values $\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \simeq 0.04444, 3.75$ and 4.66667 are away from the interval $[(1-\sqrt{c})^2, (1+\sqrt{c})^2] \simeq [0.08578, 2.91422]$. The histogram and the scatterplot of Figure 1 is from Gaussian samples when $p = 1000, n = 2000$. The smooth curve is the theoretical limiting density and the theoretical locations of the three separated eigenvalues are plotted with + signs below the horizontal axis. The smallest and largest two sample eigenvalues are plotted with + signs about the horizontal axis. Figure 2 is from Gaussian samples when $p = 100, n = 200$ while Figures 3 and 4 from samples of Bernoulli variables taking values -1 or 1 when $p = 1000, n = 2000$, and $p = 100, n = 200$ respectively. The observed values of the four separated eigenvalues in each case are as follows:

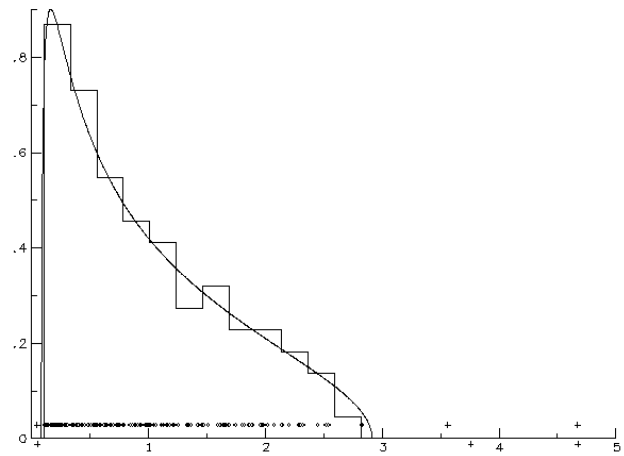

Figure 2: Gaussian samples when $p = 100, n = 200$

	smallest eigenvalue	2nd largest eigenvalue	largest eigenvalue
theoretical	0.04444	3.75	4.66667
Gaussian $p = 1000$	0.04369	3.78400	4.59127
Gaussian $p = 100$	0.03979	3.55388	4.66192
Bernoulli $p = 1000$	0.04555	3.75706	4.66594
Bernoulli $p = 100$	0.05015	3.62337	4.70786

Figure 5 and Figure 6 are the cases when $c = 2$, $p = 2000, n = 1000$ with Gaussian and Bernoulli samples, respectively. Again three non-unit population eigenvalues are chosen 0.1, 3 and 4. The critical value of the eigenvalues is $1 + \sqrt{c} \simeq 2.41421$ and the theory predicts that the two largest sample eigenvalues given by $\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \simeq 6$ and 6.66667 are separated from the interval $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2] \simeq [0.17157, 3.41209]$. Only non-zero eigenvalues are plotted in Figure 5 and Figure 6. The observed values of the separated eigenvalues in each case are as follows:

	2nd largest eigenvalue	largest eigenvalue
theoretical	6	6.66667
Gaussian $p = 2000$	5.8523	6.4013
Bernoulli $p = 2000$	6.01065	6.725

The paper is organized as follows. In section 2, we summarize the work of Z. D. Bai and J. W. Silverstein on which we heavily rely to prove our results. It turns out that the determination of the support of a Stieltjes transform plays the crucial role. This is obtained in section 3. The proofs of the main theorems are given in section 4.

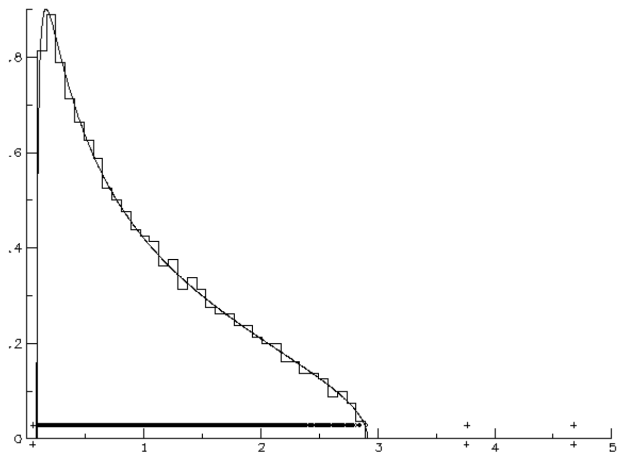

Figure 3: Bernoulli samples taking values -1 or 1 when $p = 1000, n = 2000$

Acknowledgments. Special thanks are due to Min Kang for kindly inviting J.B. to give a talk at North Carolina State University where the authors happened to have a chance to discuss about the problem, which eventually lead to this work. We would also like to thank Iain Johnstone for telling us the work of Debashis Paul [16] which was being done independently and at the same time. The work of J.B. was supported in part by NSF Grant #DMS-0350729.

2 Results of Z. D. Bai and J. W. Silverstein

Our analysis replies heavily on the work [2, 3] of Bai and Silverstein. In this section, we summarize the necessary results from [2, 3].

Notational Remark. We denote by p the population size and by n the sample size. The notations n and N are used in [3] for p and n, respectively.

For a distribution function $G(\lambda)$, its Stieltjes transform $m_G(z)$ is defined by

$$m_G(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dG(\lambda), \qquad z \in \mathbb{C}^+ := \{z \in \mathbb{C} : Im(z) > 0\}. \tag{2.1}$$

Also recall the inversion formula

$$G([a,b]) = \frac{1}{\pi} \lim_{\eta \downarrow 0} \int_a^b Im\big(m_G(\xi + i\eta)\big) d\xi \tag{2.2}$$

for continuity points a, b of G.

Assume the following:

(a) Z_{ij} are i.i.d. random variables in $\mathbb{C}$ with $\mathbb{E}(Z_{11}) = 0$, $\mathbb{E}|Z_{11}|^2 = 1$ and $\mathbb{E}|Z_{11}|^4 < \infty$.

(b) $n = n(p)$ with $c_p := \frac{p}{n} \to c > 0$ as $p \to \infty$.

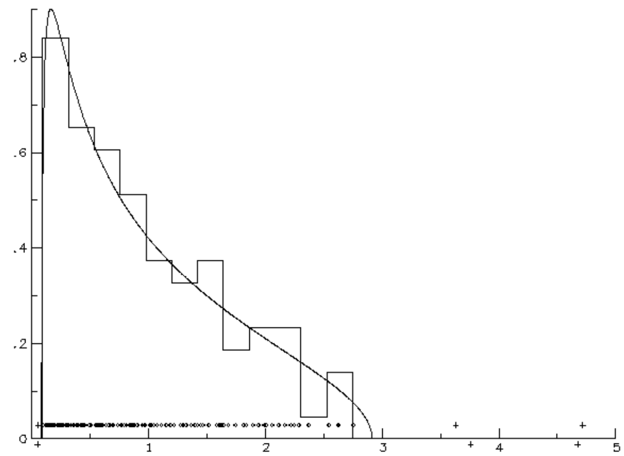

Figure 4: Bernoulli samples taking values -1 or 1 when $p = 100, n = 200$

(c) For each p, $U_T T_p U_T^{-1} = diag(\sigma_1^{(p)}, \ldots, \sigma_p^{(p)})$ for some unitary matrix U_T such that $H_p \to H_\infty$ in distribution for some distribution function H_∞ where H_p is the empirical distribution function of the eigenvalues of T_p defined by

$$dH_p(\lambda) = \frac{1}{p} \sum_{j=1}^{p} \delta_{\sigma_j^{(p)}}(\lambda). \tag{2.3}$$

(d) $\max\{\sigma_1^{(p)}, \ldots, \sigma_p^{(p)}\}$ is bounded in p.

(e) Set $Z_p = (Z_{ij})$, $1 \le i \le p$, $1 \le j \le n$ and $B_p = \frac{1}{n} T_p^{1/2} Z_p Z_p^* T_p^{1/2}$.

(f) Set

$$z_p(m) = -\frac{1}{m} + c_p \int \frac{t}{1+tm} dH_p(t). \tag{2.4}$$

¿From [22] and [21], it is known that there is a unique inverse function $m_p(z)$ such that $m_p(z) \in \mathbb{C}^+$ for $z \in \mathbb{C}^+$ It is also known [22, 21] that $m_p(z)$ is the Stieltjes transform of a distribution, which will be denoted by F_p:

$$m_p(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF_p(\lambda), \qquad z \in \mathbb{C}^+. \tag{2.5}$$

Suppose that the interval $[a, b]$ with $a > 0$ lies in an open interval outside the support of F_p for all large p.

Now we will state the main result of [3] which we need for our analysis. It is easy to check that F_p converges to some distribution function F_∞. Then [21] $F_\infty(\lambda)$ is the almost sure limit of the empirical spectral distribution of $\underline{B}_p := \frac{1}{n} Z_p^* T_p Z_p$ and

$$F(\lambda) := \frac{1}{c}\big(F_\infty(\lambda) - (1-c)1_{[0,\infty)}\big) \tag{2.6}$$

is the almost sure limit of the empirical spectral distribution of $B_p = \frac{1}{n} T_p^{1/2} Z_p Z_p^* T_p^{1/2}$.

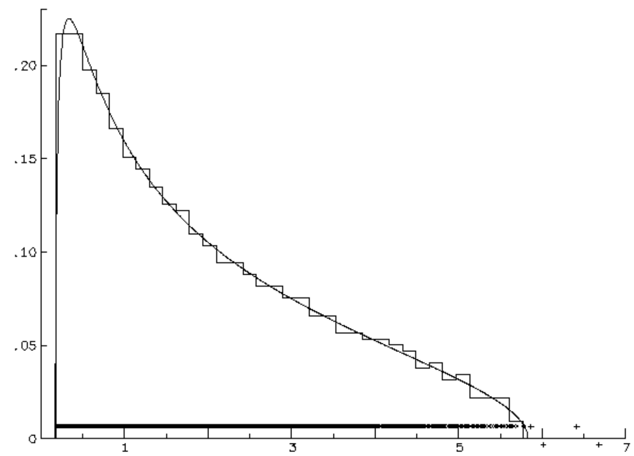

Figure 5: Gaussian samples when $p = 2000, n = 1000$

Remark. The function F_p is not the empirical distribution of B_p. The distribution function F_p is defined only through (2.5).

Moreover [2], the Stieltjes transform of F_∞,

$$m_\infty(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF_\infty(\lambda), \qquad z \in \mathbb{C}_+ := \{z \in \mathbb{C} : Im(z) > 0\}, \tag{2.7}$$

is invertible, with the inverse given by

$$z_\infty(m) = -\frac{1}{m} + c \int \frac{t}{1+tm} dH_\infty(t). \tag{2.8}$$

On the other hand, given H_∞, F_∞ is determined from (2.8) and (2.7). Note that $m_\infty(z)$ is well-defined not only on $\mathbb{C}_+$ but also up to the real line outside $\mathrm{supp}(F_\infty)$ and its inverse exists on $m_\infty(\mathbb{C}_+ \cup \mathrm{supp}(F_\infty)^c)$.

Remark. If $[a, b]$ satisfies condition (f) above, it is easy to check that $[a, b] \subset \mathrm{supp}(F_\infty)^c$.

Given an interval $[a, b]$ satisfying condition (f) above and $m_\infty(b) < 0$, it is shown in [3] that there is an integer $i_p \geq 0$ satisfying the conditions

$$\sigma_{i_p}^{(p)} > -\frac{1}{m_\infty(b)}, \qquad \sigma_{i_p+1}^{(p)} < -\frac{1}{m_\infty(a)} \tag{2.9}$$

for large p. (Here $\sigma_0^{(p)} := \infty$.)

Proposition 2.1 (Theorem 1.2 [3]). *Assume (a)-(f) above.*

(i) If $c(1 - H_\infty(0)) > 1$, then x_0, the smallest value in the support of F_∞, is positive, and $s_n^{(p)} \to x_0$ with probability 1. The value x_0 is the maximum of the function $z_\infty(m)$ for $m \in \mathbb{R}_+$.

(ii) If $c(1 - H_\infty(0)) \leq 1$ or $c(1 - H_\infty(0)) > 1$ but $[a, b]$ is not contained in $[0, x_0]$, then $m_\infty(b) < 0$ and

$$\mathbb{P}\big(s_{i_p}^{(p)} > b \text{ and } s_{i_p+1}^{(p)} < a \text{ for all large } p\big) = 1 \tag{2.10}$$

with i_p defined in (2.9). (Here $s_0^{(p)} := \infty$.)

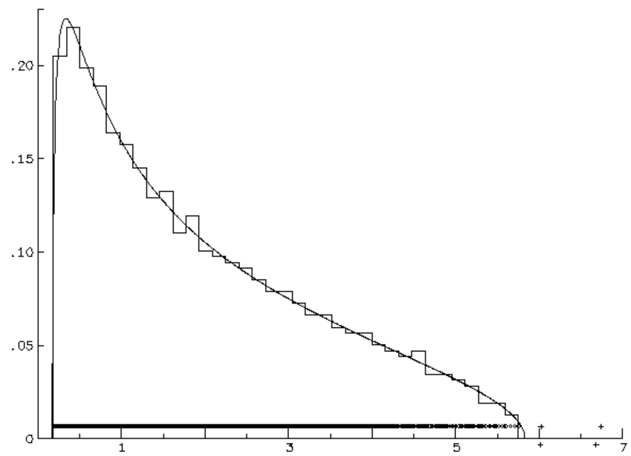

Figure 6: Bernoulli samples taking values -1 or 1 when $p = 2000, n = 1000$

3 Determination of $\mathrm{supp}(F_p)$

The key part in applying Proposition 2.1 turns out to be determining the support of F_p. This can be extracted from the following result due to Silverstein and Choi.

Lemma 3.1 ([23]; see also Lemma 1.3 [3]). *If $x \notin \mathrm{supp}(F_p)$, then $m := m_p(x)$ satisfies*

(i) $m \in \mathbb{R} \setminus \{0\}$

(ii) $-\frac{1}{m} \notin \mathrm{supp}(H_p)$

(iii) $z_p'(m) > 0$.

Conversely, if m satisfies (i)-(iii), then $x = z_p(m) \notin \mathrm{supp}(F_p)$.

Remark. Lemma 1.3 of [3] is stated for H_∞. But the proof of Lemma 1.3 in [23] applies also to the finite p case of H_p without any change. Indeed, the proposition applies to any distribution defined by its Stieltjes transform as in (2.5).

Remark. It is also shown in [23] that F_p has continuous density on $\mathbb{R}_+$.

When T_p is as in (1.8),

$$dH_p(x) = \frac{1}{p}\sum_{j=1}^{M} k_j \delta_{\alpha_j}(x) + \left(1 - \frac{r}{p}\right)\delta_1(x) \tag{3.1}$$

and

$$z_p(m) = -\frac{1}{m} + \frac{c_p}{1+m} + \frac{1}{n}\left[\sum_{j=1}^{M} \frac{k_j \alpha_j}{1+\alpha_j m} - \frac{r}{1+m}\right], \tag{3.2}$$

where we recall that $r = k_1 + \cdots + k_M$. We first determine the set of real m such that $z_p'(m) > 0$.

Now

$$
\begin{aligned}
z_p'(m) &= \frac{1}{m^2} - \frac{c_p}{(1+m)^2} + \frac{1}{n}\left[\sum_{j=1}^{M} \frac{-k_j\alpha_j^2}{(1+\alpha_j m)^2} + \frac{r}{(1+m)^2}\right] \\
&= \frac{f(m) + \frac{1}{n}g(m)}{m^2(1+m)^2\prod_{\ell=1}^{M}(1+\alpha_\ell m)^2},
\end{aligned}
\tag{3.3}
$$

where

$$
f(m) := \left((1+m)^2 - c_p m^2\right)\prod_{\ell=1}^{M}(1+\alpha_\ell m)^2 \tag{3.4}
$$

and

$$
g(m) := \left[\sum_{j=1}^{M} \frac{-k_j\alpha_j^2}{(1+\alpha_j m)^2} + \frac{r}{(1+m)^2}\right] m^2(1+m)^2\prod_{\ell=1}^{M}(1+\alpha_\ell m)^2. \tag{3.5}
$$

We need the following basic lemmas of complex variables to determine the solution of $z_p'(m) = 0$.

Lemma 3.2. *Let $h(z)$ be an analytic function in a closed disk $\overline{D(z_0, r)}$ of radius $r > 0$ centered at z_0. Then there is $\epsilon_0 > 0$ such that for $0 \le \epsilon < \epsilon_0$, the equation*

$$
z - z_0 = \epsilon h(z) \tag{3.6}
$$

has a unique solution in $D(z_0, r)$, which satisfies

$$
z = z_0 + \epsilon h(z_0) + O(\epsilon^2). \tag{3.7}
$$

Furthermore, if z_0 is real and $h(z)$ is real for real z, the solution (3.7) *is real.*

Proof. As h is continuous, there is a constant $C > 0$ such that $|h(z)| \le C$ for $|z - z_0| \le r$. When $|\epsilon| < \frac{r}{C}$, for $|z - z_0| = r$,

$$
|z - z_0| = r > |\epsilon| C \ge |\epsilon h(z)|. \tag{3.8}
$$

Hence from Rouche's theorem, the number of zeros of $z - z_0 - \epsilon h(z)$ inside $D(z_0, r)$ is equal to the number of zeros of $z - z_0$ inside $D(z_0, r)$, which is one. The zero z_ϵ satisfies $z_\epsilon - z_0 = \epsilon h(z_\epsilon) = O(\epsilon)$. Thus

$$
z_\epsilon - z_0 - \epsilon h(z_0) = \epsilon(h(z_\epsilon) - h(z_0)) = O(\epsilon^2). \tag{3.9}
$$

If $h(z)$ is real for real z, then by taking complex conjugate of (3.6), we find that $\overline{z_\epsilon}$ is also a solution. Since there is only one solution, we find that z_ϵ is real. $\square$

Lemma 3.3. *Let $h(z)$ be an analytic function in a closed disk $\overline{D(z_0, r)}$ of radius $r > 0$ centered at z_0 such that $h(z_0) \neq 0$. Then there are $0 < r_0 \le r$ and $\epsilon_0 > 0$ such that for $0 \le \epsilon < \epsilon_0$, the equation*

$$
(z - z_0)^2 = \epsilon h(z) \tag{3.10}
$$

have precisely two distinct solutions in $D(z_0, r_0)$, which satisfy

$$
z = z_0 \pm \sqrt{\epsilon}\sqrt{h(z_0)} + O(\epsilon) \tag{3.11}
$$

where $\sqrt{h(z_0)}$ is an arbitrary branch. Furthermore, suppose that z_0 is real and $h(z)$ is real for real z. Then if $h(z_0) > 0$, both solutions (3.11) *are real. On the other hand, if $h(z_0) < 0$, both solutions* (3.11) *are non-real.*

Proof. The proof of (3.11) follows from Lemma 3.2 by taking the square root of (3.10). When z_0 is real and $h(z)$ is real for real z, the complex conjugate of a solution of (3.10) is also a solution. Thus the two solutions(3.11) of (3.10) are either complex conjugates of each other or both real since there are precisely two distinct solutions. Hence the Lemma follows. $\square$

For the remainder of this section, we assume that $c \neq 1$ and none of α_j's are equal to $1 \pm \sqrt{c}$. We further assume that p and n are sufficiently large so that $c_p \neq 1$ and none of α_j's are equal to $1 \pm \sqrt{c_p}$. Then the numerator of (3.3) is a polynomial of degree exactly $2M + 2$, and we now determine all the solutions of $z_p'(m) = 0$.

For f defined in (3.4), the equation $f(m) = 0$ has distinct solutions

$$m = \frac{-1}{1+\sqrt{c_p}} =: m_+, \qquad m = \frac{-1}{1-\sqrt{c_p}} =: m_- \tag{3.12}$$

of multiplicity 1 and

$$m = \frac{-1}{\alpha_j}, \qquad j = 1, 2, \ldots, M, \tag{3.13}$$

of multiplicity 2. The roots of $z_p'(m)$ are expected to be perturbations of the roots of $f(m)$, which we will find. First consider m_+. Dividing the equation $f(m) + \frac{1}{n}g(m) = 0$ by $\frac{f(m)}{m-m_+}$, we obtain the equation

$$m - m_+ + \frac{1}{n}\frac{m^2(1+m)^2}{(1-c_p)(m-m_-)}\left[\sum_{j=1}^{M}\frac{-k_j\alpha_j^2}{(1+\alpha_j m)^2} + \frac{r}{(1+m)^2}\right] = 0. \tag{3.14}$$

Lemma 3.2 implies that there is a solution of $z_p'(m) = 0$ of the form

$$m = m_+ + O\left(\frac{1}{n}\right), \tag{3.15}$$

which is real. Similarly, there is a real solution of $z_p'(m) = 0$ of the form

$$m = m_- + O\left(\frac{1}{n}\right). \tag{3.16}$$

Now consider the root $m = \frac{-1}{\alpha_j}$ of $f(m) = 0$. Dividing $f + \frac{1}{n}g = 0$ by $\frac{f(m)}{(m+\frac{1}{\alpha_j})^2}$, we obtain the equation

$$\left(m + \frac{1}{\alpha_j}\right)^2 = \frac{1}{n}G_j(m) \tag{3.17}$$

where

$$G_j(m) = \frac{-(1+\alpha_j m)^2 m^2 (1+m)^2}{\alpha_j^2(1-c_p)(m-m_+)(m-m_-)}\left[\sum_{\ell=1}^{M}\frac{-k_\ell\alpha_\ell^2}{(1+\alpha_\ell m)^2} + \frac{r}{(1+m)^2}\right]. \tag{3.18}$$

Note that

$$G_j(-\frac{1}{\alpha_j}) = \frac{k_j(\alpha_j-1)^2}{\alpha_j^4(1-c_p)(\frac{-1}{\alpha_j}-m_+)(\frac{-1}{\alpha_j}-m_-)} \tag{3.19}$$

is not zero and also $G_j(m)$ is real for real m. Thus Lemma 3.3 implies that there are precisely two solutions of $z_p'(m) = 0$ of the form

$$m = -\frac{1}{\alpha_j} \pm \frac{1}{\sqrt{n}}\sqrt{G_j(-\frac{1}{\alpha_j})} + O\left(\frac{1}{n}\right), \qquad j = 1, \ldots, M, \tag{3.20}$$

where the pair for each j are either both real or both non-real depending on the sign of $G_j\left(-\frac{1}{\alpha_j}\right)$.

Now when $c_p < 1$, the condition $G_j\left(-\frac{1}{\alpha_j}\right) > 0$ is equivalent to

$$\frac{-1}{\alpha_j} > m_+ \quad \text{or} \quad \frac{-1}{\alpha_j} < m_-, \tag{3.21}$$

which is the same as

$$\alpha_j > 1 + \sqrt{c_p} \quad \text{or} \quad \alpha_j < 1 - \sqrt{c_p}. \tag{3.22}$$

On the other hand, when $c_p > 1$, we note that $m_+ < 0 < m_-$. The condition $G_j\left(-\frac{1}{\alpha_j}\right) > 0$ is now equivalent to

$$m_+ < \frac{-1}{\alpha_j} < m_-, \tag{3.23}$$

which is the same as (since $\alpha_j > 0$)

$$\alpha_j > 1 + \sqrt{c_p}. \tag{3.24}$$

We summarize the above calculations.

Lemma 3.4. *The solutions of* $z_p'(m) = 0$ *are*

$$m = -\frac{1}{1+\sqrt{c_p}} + O\left(\frac{1}{n}\right) =: m_+^{(n)}, \qquad m = -\frac{1}{1-\sqrt{c_p}} + O\left(\frac{1}{n}\right) =: m_-^{(n)}. \tag{3.25}$$

and

$$m = -\frac{1}{\alpha_j} \pm \frac{1}{\sqrt{n}}\sqrt{G_j\left(-\frac{1}{\alpha_j}\right)} + O\left(\frac{1}{n}\right) =: m_{j,\pm}^{(n)}, \qquad j = 1, \ldots, M, \tag{3.26}$$

all of multiplicity 1. *Furthermore, the following holds.*

- *When* $c_p < 1$, $m_-^{(n)} < m_+^{(n)} < 0$, *and* $m_{j,\pm}^{(n)}$ *are real if and only if* $\alpha_j > 1 + \sqrt{c_p}$ *or* $\alpha_j < 1 - \sqrt{c_p}$. *If* $1 - \sqrt{c_p} < \alpha_j < 1 + \sqrt{c_p}$, $m_{j,+}^{(n)}$ *and* $m_{j,-}^{(n)}$ *are complex conjugates of each other.*
- *When* $c_p > 1$, $m_+^{(n)} < 0 < m_-^{(n)}$, *and* $m_{j,\pm}^{(n)}$ *are real if and only if* $\alpha_j > 1 + \sqrt{c_p}$. *If* $\alpha_j < 1 + \sqrt{c_p}$, $m_{j,+}^{(n)}$ *and* $m_{j,-}^{(n)}$ *are complex conjugates of each other.*

We now consider the cases when $c < 1$ and when $c > 1$ separately.

3.1 When $c < 1$

Let the indices $0 \le M_0, M_1 \le M$ be defined as in Theorem 1.1 (recall that we assume that none of the α_j's are equal to $1 \pm \sqrt{c}$), so that

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{c} > \alpha_{M_0+1} > \cdots > \alpha_{M-M_1} > 1 - \sqrt{c} > \alpha_{M-M_1+1} > \cdots > \alpha_M. \tag{3.27}$$

We now find the intervals in which $z_p'(m) > 0$.

The denominator of (3.3) is non-negative. ¿From Lemma 3.4, the numerator of (3.3) is factored as

$$const \cdot (m - m_-^{(n)})(m - m_+^{(n)}) \prod_{j=1}^{M} (m - m_{j,-}^{(n)})(m - m_{j,+}^{(n)}) \tag{3.28}$$

The constant prefactor is, from (3.4) and (3.5),

$$(1-c_p)\prod_{j=1}^{M}\alpha_j^2+O\Big(\frac{1}{n}\Big), \tag{3.29}$$

which is positive when n is large enough. On the other hand, among the terms in the product of (3.28), $m_{j,\pm}^{(n)}$ corresponding to the indices $M_0+1\le j\le M_1$ are complex conjugates of each other. Thus

$$\prod_{j=M_0+1}^{M_1+1}(m-m_{j,-}^{(n)})(m-m_{j,+}^{(n)})\ge 0. \tag{3.30}$$

Hence using the fact that

$$0>m_{1,+}^{(n)}>m_{1,-}^{(n)}>\cdots>m_{M_0,+}^{(n)}>m_{M_0,-}^{(n)}>m_+^{(n)} \tag{3.31}$$

and

$$m_-^{(n)}>m_{M_1+1,+}^{(n)}>m_{M_1+1,-}^{(n)}>\cdots>m_{M,+}^{(n)}>m_{M,-}^{(n)}, \tag{3.32}$$

we find that the numerator of (3.3) is positve in the intervals

$$(-\infty,m_{M,-}^{(n)})\cup(m_{M,+}^{(n)},m_{M-1,-}^{(n)})\cup\cdots\cup(m_{M_1+2,+}^{(n)},m_{M_1+1,-}^{(n)})\cup(m_{M_1+1,+}^{(n)},m_-^{(n)}) \tag{3.33}$$

union

$$(m_+^{(n)},m_{M_0,-}^{(n)})\cup(m_{M_0,+}^{(n)},m_{M_0-1,-}^{(n)})\cup\cdots\cup(m_{2,+}^{(n)},m_{1,-}^{(n)})\cup(m_{1,+}^{(n)},\infty). \tag{3.34}$$

The singular points of (3.3) are not contained in any of the above intervals except for the singular point $m=0$. Hence the set of m such that $z_p'(m)>0$ is equal to (3.33) union

$$(m_+^{(n)},m_{M_0,-}^{(n)})\cup(m_{M_0,+}^{(n)},m_{M_0-1,-}^{(n)})\cup\cdots\cup(m_{2,+}^{(n)},m_{1,-}^{(n)})\cup(m_{1,+}^{(n)},0)\cup(0,\infty). \tag{3.35}$$

Now Lemma 3.1 determines $\operatorname{supp}(F_p)$.

Proposition 3.5. *Suppose that $c<1$ and none of α_j is equal to $1\pm\sqrt{c}$. With the indices M_0 and M_1 defined in Theorem 1.1, for n sufficiently large,*

$$\begin{aligned}\operatorname{supp}(F_p)^c=&(-\infty,0)\cup(0,z_{M,-}^{(n)})\cup(z_{M,+}^{(n)},z_{M-1,-}^{(n)})\cup\cdots\cup(z_{M_1+1,+}^{(n)},z_-^{(n)})\\&\cup(z_+^{(n)},z_{M_0,-}^{(n)})\cup(z_{M_0,+}^{(n)},z_{M_0-1,-}^{(n)})\cup\cdots\cup(z_{2,+}^{(n)},z_{1,-}^{(n)})\cup(z_{1,+}^{(n)},\infty)\end{aligned} \tag{3.36}$$

where

$$z_\pm^{(n)}=(1\pm\sqrt{c_p})^2+O\Big(\frac{1}{n}\Big) \tag{3.37}$$

and

$$z_{j,\pm}^{(n)}=\alpha_j+\frac{c_p\alpha_j}{\alpha_j-1}\pm\frac{A_j}{\sqrt{n}}+O\Big(\frac{1}{n}\Big),\qquad j=1,\dots,M_0,\quad j=M_1+1,\dots,M, \tag{3.38}$$

for some constant $A_j>0$. The intervals in (3.36) *are disjoint.*

Proof. We will first see that the intervals (3.33) union (3.35) satisfy the conditions (i)-(iii) of Lemma 3.1. The condition (iii) is clearly satisfied. Also 0 is not contained in (3.33) and (3.35), and so condition (i) is fullfilled. Finally, as $\mathrm{supp}(H_p) = \{\alpha_1, \dots, \alpha_M, 1\}$ and

$$m_-^{(n)} < -1 < m_+^{(n)}, \qquad m_{j,-}^{(n)} < -\frac{1}{\alpha_j} < m_{j,+}^{(n)}, \tag{3.39}$$

the condition (ii) is satisfied for m in (3.33) union (3.35).

We now need to find the image of the above intervals under z_p. Clearly, $z_p(-\infty) = 0$, $z_p(0-) = +\infty$, $z_p(0+) = -\infty$ and $z_p(+\infty) = 0$. A direct computation yields

$$z_p(m_\pm^{(n)}) = (1 \pm \sqrt{c_p})^2 + O\left(\frac{1}{n}\right). \tag{3.40}$$

and

$$z_p(m_{j,\pm}^{(n)}) = \alpha_j + \frac{c_p \alpha_j}{\alpha_j - 1} \pm \frac{A_j}{\sqrt{n}} + O\left(\frac{1}{n}\right) \tag{3.41}$$

where

$$A_j = \frac{1}{C_j}\left\{C_j^2 \alpha_j^2 \left(1 - \frac{c_p}{(\alpha_j - 1)^2}\right) + k_j\right\}, \qquad C_j := \sqrt{G(-1/\alpha_j)}. \tag{3.42}$$

Note that $A_j > 0$ for $1 \le j \le M_0$ and $M_1 + 1 \le j \le M$ since $\alpha_j > 1 + \sqrt{c_p}$ or $\alpha_j < 1 - \sqrt{c_p}$. Also it is straightforward to check from the graph of the function

$$x + \frac{c_p x}{x - 1} \tag{3.43}$$

that

$$\begin{aligned} 0 < \alpha_M + \frac{c_p \alpha_M}{\alpha_M - 1} < \cdots < \alpha_{M_1+1} + \frac{c_p \alpha_{M_1+1}}{\alpha_{M_1+1} - 1} < (1 - \sqrt{c_p})^2 \\ < (1 + \sqrt{c_p})^2 < \alpha_{M_0} + \frac{c_p \alpha_{M_0}}{\alpha_{M_0} - 1} < \cdots < \alpha_1 + \frac{c_p \alpha_1}{\alpha_1 - 1}. \end{aligned} \tag{3.44}$$

This implies the Proposition. □

3.2 When $c > 1$

This case is similar to the previous case when $c < 1$. We indicate only the difference.

We again assume that p and n are large enough so that the set of j's satisfying $\alpha_j > 1 + \sqrt{c_p}$ is the same as the set of j's satisfying $\alpha_j > 1 + \sqrt{c}$. Let the index $0 \le M_0 \le M$ be defined, as in Theorem 1.1. We further assume that none of α_j is equal to $1 + \sqrt{c}$ so that

$$\alpha_{M_0} > 1 + \sqrt{c} > \alpha_{M_0+1}. \tag{3.45}$$

The denominator of (3.3) is non-negative and as before, the numerator of (3.3) is equal to (3.28). But this time, the constant prefactor (3.29) is negative when n is large enough. Also as in (3.30),

$$\prod_{j=M_0+1}^{M} (m - m_{j,-}^{(n)})(m - m_{j,+}^{(n)}) \ge 0. \tag{3.46}$$

Now using the fact that

$$m_-^{(n)} > 0 > m_{1,+}^{(n)} > m_{1,-}^{(n)} > \cdots > m_{M_0,+}^{(n)} > m_{M_0,-}^{(n)} > m_+^{(n)}, \tag{3.47}$$

we find that the numerator of (3.3) is positive in the intervals

$$(m_+^{(n)}, m_{M_0,-}^{(n)}) \cup (m_{M_0,+}^{(n)}, m_{M_0-1,-}^{(n)}) \cup \cdots \cup (m_{2,+}^{(n)}, m_{1,-}^{(n)}) \cup (m_{1,+}^{(n)}, m_-^{(n)}). \tag{3.48}$$

Hence taking into accounts of the singular point $m = 0$ of $z_p'(m)$, the intervals where $z_p'(m) > 0$ is

$$(m_+^{(n)}, m_{M_0,-}^{(n)}) \cup (m_{M_0,+}^{(n)}, m_{M_0-1,-}^{(n)}) \cup \cdots \cup (m_{2,+}^{(n)}, m_{1,-}^{(n)}) \cup (m_{1,+}^{(n)}, 0) \cup (0, m_-^{(n)}). \tag{3.49}$$

The proof of the following proposition is parallel to Proposition 3.5.

Proposition 3.6. *Suppose that $c > 1$ and none of α_j is equal to $1 + \sqrt{c}$. With the index M_0 defined in Theorem 1.2, for n sufficiently large,*

$$\mathrm{supp}(F_p)^c = (-\infty, z_-^{(n)}) \cup (z_+^{(n)}, z_{M_0,-}^{(n)}) \cup (z_{M_0,+}^{(n)}, z_{M_0-1,-}^{(n)}) \cup \cdots \cup (z_{2,+}^{(n)}, z_{1,-}^{(n)}) \cup (z_{1,+}^{(n)}, \infty) \tag{3.50}$$

where

$$z_\pm^{(n)} = (1 \pm \sqrt{c_p})^2 + O\left(\frac{1}{n}\right) \tag{3.51}$$

and

$$z_{j,\pm}^{(n)} = \alpha_j + \frac{c_p \alpha_j}{\alpha_j - 1} \pm \frac{A_j}{\sqrt{n}} + O\left(\frac{1}{n}\right), \qquad j = 1, \ldots, M_0, \tag{3.52}$$

for some constant $A_j > 0$. The intervals in (3.50) are disjoint.

4 Proof of Theorems 1.1, 1.2 and 1.3

When T_p is (1.8), as H_p is equal to (3.1),

$$dH_\infty(x) = \delta_1(x). \tag{4.1}$$

Hence

$$z_\infty(m) = -\frac{1}{m} + c\int \frac{t}{1+tm} dH_\infty(t) = -\frac{1}{m} + \frac{c}{1+m}. \tag{4.2}$$

It is well-known that ([22], see also Theorem 3.4 of [1]) in this case,

$$dF_\infty(\lambda) = \begin{cases} \frac{1}{2\pi\lambda}\sqrt{((1+\sqrt{c})^2 - \lambda)(\lambda - (1-\sqrt{c})^2)}1_{[(1-\sqrt{c})^2, (1+\sqrt{c})^2]}(\lambda), & c > 1 \\ \frac{1}{2\pi\lambda}\sqrt{((1+\sqrt{c})^2 - \lambda)(\lambda - (1-\sqrt{c})^2)}1_{[(1-\sqrt{c})^2, (1+\sqrt{c})^2]}(\lambda) + (1-c)\delta_0, & 0 < c \le 1. \end{cases} \tag{4.3}$$

4.1 When $c < 1$

We first assume that none of α_j is equal to $1 \pm \sqrt{c}$ so that Proposition 3.5 is applicable. The case when some of α_j are equal to $1 \pm \sqrt{c}$ will be discussed at the end of this subsection.

When T_p is (1.8), all the conditions (a)-(e) of Proposition 2.1 are satisfied or are defined accordingly.

Now suppose $[a,b]$ is an interval satisfying condition (f). Since

$$z_+^{(n)} \to (1+\sqrt{c})^2, \qquad z_-^{(n)} \to (1-\sqrt{c})^2, \tag{4.4}$$

and for any i,

$$z_{i,+}^{(n)}, \quad z_{i,-}^{(n)} \to \alpha_i + \frac{c\alpha_i}{\alpha_i - 1}, \tag{4.5}$$

we see that

$$\begin{aligned}[a,b] \subset & \big(-\infty, 0\big) \cup \left(0, \alpha_M + \frac{c\alpha_M}{\alpha_M - 1}\right) \cup (\alpha_M + \frac{c\alpha_M}{\alpha_M - 1}, \alpha_{M-1} + \frac{c\alpha_{M-1}}{\alpha_{M-1}-1}\Big) \\ & \cup \cdots \cup \left(\alpha_{M_1+1} + \frac{c\alpha_{M_1+1}}{\alpha_{M_1+1}-1}, (1-\sqrt{c})^2\right) \\ & \cup \left((1+\sqrt{c})^2, \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0}-1}\right) \cup \cdots \cup \left(\alpha_2 + \frac{c\alpha_2}{\alpha_2 - 1}, \alpha_1 + \frac{c\alpha_1}{\alpha_1 - 1}\right) \cup \left(\alpha_1 + \frac{c\alpha_1}{\alpha_1-1}, \infty\right).\end{aligned} \tag{4.6}$$

On the other hand,

$$\mathrm{supp}(F_\infty)^c = (-\infty, 0) \cup \big(0, (1-\sqrt{c})^2\big) \cup \big((1+\sqrt{c})^2, \infty\big). \tag{4.7}$$

Hence $[a,b] \subset \mathrm{supp}(F_\infty)^c$. Also from definition (2.7), it is easy to see that $m'_\infty(z) > 0$ for $z \in \mathrm{supp}(F_\infty)^c$. The first consequence of (ii) of Proposition 2.1 (note that $H_\infty(0) = 0$) is that $m_\infty(b) < 0$. Thus $m_\infty(a) < m_\infty(b) < 0$. Therefore, the condition (2.9) is equivalent to the condition

$$[a,b] \subset \big[z_\infty\big(-1/\sigma_{i_p+1}^{(p)}\big), z_\infty\big(-1/\sigma_{i_p}^{(p)}\big)\big]. \tag{4.8}$$

We will consider four different choices of $[a,b]$. First fix $1 \le j \le M_0$. Take

$$[a,b] = \big[\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} + \epsilon, \alpha_{j-1} + \frac{c\alpha_{j-1}}{\alpha_{j-1}-1} - \epsilon\big] \tag{4.9}$$

for an arbitrary fixed $\epsilon > 0$. (Here $\alpha_0 := +\infty$.) From (4.5), we see that

$$[a,b] \subset (z_{j,+}^{(n)}, z_{j-1,-}^{(n)}) \tag{4.10}$$

for all large p, and hence condition (f) is satisfied using Proposition 3.5. Set

$$i_p := k_1 + \cdots + k_{j-1}. \tag{4.11}$$

(When $j = 1$, $i_p := 0$.) For T_p given by (1.8),

$$\sigma_{i_p}^{(p)} = \alpha_{j-1}, \qquad \sigma_{i_p+1}^{(p)} = \alpha_j. \tag{4.12}$$

But

$$z_\infty(-1/\alpha_j) = \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \tag{4.13}$$

and hence the condition (4.8) is satisfied. Therefore i_p is defined to satisfy the condition (2.9). Proposition 3.5 now implies that

$$\mathbb{P}\left(s_{k_1+\cdots+k_{j-1}}^{(p)} > \alpha_{j-1} + \frac{c\alpha_{j-1}}{\alpha_{j-1}-1} - \epsilon \text{ and } s_{k_1+\cdots+k_{j-1}+1}^{(p)} < \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} + \epsilon \text{ for all large } p\right) = 1. \tag{4.14}$$

This yields that, $1 \le j \le M_0 - 1$,

$$\mathbb{P}\bigg(\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} - \epsilon < s^{(p)}_{k_1+\cdots+k_{j-1}+k_j} \le \cdots \le s^{(p)}_{k_1+\ldots k_{j-1}+1} < \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} + \epsilon \text{ for all large } p\bigg) = 1, \quad (4.15)$$

which implies (1.10) for $1 \le j \le M_0 - 1$, and

$$\mathbb{P}\bigg(s^{(p)}_{k_1+\cdots+k_{M_0-1}+1} < \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0} - 1} + \epsilon \text{ for all large } p\bigg) = 1. \quad (4.16)$$

For the second choice of $[a, b]$, set

$$[a, b] = \big[(1 + \sqrt{c})^2 + \epsilon, \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0} - 1} - \epsilon\big] \quad (4.17)$$

for an arbitrary fixed $\epsilon > 0$. Noting that

$$z^{(n)}_+ \to (1 + \sqrt{c})^2 \quad (4.18)$$

and setting $i_p := k_1 + \cdots + k_{M_0}$, a calculation similar to the above yields that

$$\mathbb{P}\bigg(s^{(p)}_{k_1+\cdots+k_{M_0}} > \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0} - 1} - \epsilon \text{ and } s^{(p)}_{k_1+\cdots+k_{M_0}+1} < (1 + \sqrt{c})^2 + \epsilon \text{ for all large } p\bigg) = 1. \quad (4.19)$$

Thus, together with (4.16), we obtain (1.10) for $j = M_0$. Also as (4.2) and discussions around (2.6) implies that the support of the limiting spectral distribution of B_p is $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$, we obtain (1.11).

As the third and forth choices of $[a, b]$, we set

$$[a, b] = \big[\alpha_{M_1+1} + \frac{c\alpha_{M_1+1}}{\alpha_{M_1+1} - 1} + \epsilon, (1 - \sqrt{c})^2 - \epsilon\big] \quad (4.20)$$

and

$$[a, b] = \big[\alpha_{j+1} + \frac{c\alpha_{j+1}}{\alpha_{j+1} - 1} + \epsilon, \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} - \epsilon\big] \quad (4.21)$$

for some $M_1 + 1 \le j \le M$ ($\alpha_{M+1} := 0$), respectively. Arguments as above imply the remaining part of Theorem 1.1.

We now consider the case when an α_j is equal to $1 \pm \sqrt{c}$. We first observe certain monotonicity of the eigenvalues $s^{(p)}_j$ on α_js. Note that the matrix $\underline{B}_p := \frac{1}{n} Z'_p T_p Z_p$ has the same set of eigenvalues as B_p except for $|p - n|$ zero eigenvalues. Consider a set of parameters β_j, $1 \le j \le M$, such that $\alpha_j \ge \beta_j$. Let $\hat{T}_p$ be the matrix T_p with α_j's replaced by β_j's, and set $\hat{B}_p = \frac{1}{n}\hat{T}_p^{1/2} Z_p Z'_p \hat{T}_p^{1/2}$ and $\underline{\hat{B}}_p = \frac{1}{n} Z'_p \hat{T}_p Z_p$. Then clearly, $\underline{B}_p$ and $\underline{\hat{B}}_p$ are Hermitian, and $\underline{B}_p \ge \underline{\hat{B}}_p$. Hence from the min-max principle (see e.g. [9]), we find that

$$s^{(p)}_j \ge \hat{s}^{(p)}_j \quad (4.22)$$

for all non-zero eigenvalues, where $\hat{s}^{(p)}_j$ denotes the eigenvalues of $\hat{B}_p$.

Suppose that

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{c} = \alpha_{M_0+1} > \cdots > \alpha_{M-M_1} > 1 - \sqrt{c} > \alpha_{M-M_1+1} > \cdots > \alpha_M. \quad (4.23)$$

Replacing in (4.23) α_{M_0+1} by $(1 \pm \epsilon)\alpha_{M_0+1} = (1 \pm \epsilon)(1 + \sqrt{c})$ for sufficiently small $\epsilon > 0$, the above monotonicity argument implies the following:

(i) For each $1 \le j \le M_0$,

$$\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \le \liminf s^{(p)}_{k_1+\cdots+k_{j-1}+i} \le \limsup s^{(p)}_{k_1+\cdots+k_{j-1}+i} \le \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \le i \le k_j. \quad (4.24)$$

almost surely.

(ii)

$$(1+\sqrt{c})^2 \le \liminf s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le \limsup s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le (1+\epsilon)\alpha_{M_0+1} + \frac{c(1+\epsilon)\alpha_{M_0+1}}{(1+\epsilon)\alpha_{M_0+1} - 1} \quad (4.25)$$

almost surely.

(iii)

$$(1-\sqrt{c})^2 \le \liminf s^{(p)}_{p-r+k_1+\cdots+k_{M_1}} \le \limsup s^{(p)}_{p-r+k_1+\cdots+k_{M_1}} \le (1-\sqrt{c})^2 \quad (4.26)$$

almost surely.

(iv) For each $M_1 + 1 \le j \le M$,

$$\begin{aligned} \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} &\le \liminf s^{(p)}_{p-r+k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \\ &\le \limsup s^{(p)}_{p-r+k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \le \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \qquad 1 \le i \le k_j \end{aligned} \quad (4.27)$$

almost surely.

Since

$$\lim_{\epsilon \downarrow 0} (1+\epsilon)\alpha_{M_0+1} + \frac{c(1+\epsilon)\alpha_{M_0+1}}{(1+\epsilon)\alpha_{M_0+1} - 1} = (1+\sqrt{c})^2 \quad (4.28)$$

and the above result is true for arbitrary sufficiently small $\epsilon > 0$, Theorem 1.1 follows for the case when the parameters are given by (4.23). For the case when $\alpha_{M-M_1} = 1 - \sqrt{c}$, the argument is almost identical, and we skip the details.

4.2 When $c > 1$

¿From (4.3), when $c > 1$, the smallest value in the support of F_∞ is

$$x_0 = (1-\sqrt{c})^2 > 0. \quad (4.29)$$

Hence Proposition 2.1 (i) implies that

$$s^{(p)}_n \to (1-\sqrt{c})^2. \quad (4.30)$$

Since when $p > n$, at least $p - n$ eigevalues $s^{(p)}_j$ are equal to 0, we conclude that

$$s^{(p)}_{n+1} = \cdots = s^{(p)}_p = 0. \quad (4.31)$$

Therefore, (1.19) and (1.20) are obtained.

The proof of (1.17) and (1.18) is similar to the case when $c < 1$ by using Proposition 3.6 and noting that an interval $[a, b]$ satisfying condition (f) of Proposition 2.1 is contained in

$$\big(-\infty, (1-\sqrt{c})^2\big) \cup \left((1+\sqrt{c})^2, \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0}-1}\right) \cup \cdots \cup \left(\alpha_2 + \frac{c\alpha_2}{\alpha_2 - 1}, \alpha_1 + \frac{c\alpha_1}{\alpha_1 - 1}\right) \cup \left(\alpha_1 + \frac{c\alpha_1}{\alpha_1 - 1}, \infty\right), \tag{4.32}$$

which is a subset of

$$\operatorname{supp}(F_\infty)^c = (-\infty, (1-\sqrt{c})^2\big) \cup \big((1+\sqrt{c})^2, \infty\big). \tag{4.33}$$

4.3 When $c = 1$

Since the limiting distribution (1.2) for $c = 1$ has a continuous density on the interval $(0, 4)$, it is easy to see (1.24).

We first observe a monotonicity of $s_j^{(p)}$ in n. Let $\hat{Z}_p = (Z_{ij})$, $1 \le i \le p, 1 \le j \le \hat{n}$ and let $\hat{B}_p := \frac{1}{\hat{n}} T_p^{1/2} \hat{Z}_p \hat{Z}_p' T_p^{1/2}$. When $\hat{n} > n$, it is clear that

$$\hat{n}\hat{B}_p \ge n B_p. \tag{4.34}$$

Therefore, if the ordered eigenvalues of $\hat{B}_p$ are denoted by $\hat{s}_p^{(p)}$, the min-max principle implies that

$$\hat{n}\hat{s}_j^{(p)} \ge n s_j^{(p)} \tag{4.35}$$

for all $1 \le j \le p$.

Take $\hat{n} = \left[\frac{n}{1+\epsilon}\right]$ for $\epsilon > 0$ where $[x]$ denotes the largest integer $\le x$. Then for sufficiently small $\epsilon > 0$,

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{1+\epsilon} > \alpha_{M_0+1} > \cdots > \alpha_M. \tag{4.36}$$

By applying Theorem 1.2 and using (4.35), we obtain the following:

- For each $1 \le j \le M_0$,

$$\liminf s^{(p)}_{k_1+\cdots+k_{j-1}+i} \ge \frac{1}{1+\epsilon}\left(\alpha_j + \frac{(1+\epsilon)\alpha_j}{\alpha_j - 1}\right), \qquad 1 \le i \le k_j. \tag{4.37}$$

  almost surely.

- $$\liminf s^{(p)}_{k_1+\cdots+k_{M_0}+1} \ge \frac{1}{1+\epsilon}(1+\sqrt{1+\epsilon})^2 \tag{4.38}$$

  almost surely.

On the other hand, take $\hat{n} = \left[\frac{n}{1-\epsilon}\right]$ for $\epsilon > 0$. We first assume $2 > \alpha_{M_0+1}$. Then as $\alpha_M > 0$, for sufficiently small $\epsilon > 0$,

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{1-\epsilon} > \alpha_{M_0+1} > \cdots > \alpha_M > 1 - \sqrt{1-\epsilon}, \tag{4.39}$$

and hence $M_1 = M$. By applying Theorem 1.1 and using (4.35), we obtain the following:

- For each $1 \le j \le M_0$,

$$\limsup s^{(p)}_{k_1+\cdots+k_{j-1}+i} \le \frac{1}{1-\epsilon}\left(\alpha_j + \frac{(1-\epsilon)\alpha_j}{\alpha_j - 1}\right), \qquad 1 \le i \le k_j. \tag{4.40}$$

  almost surely.

- $$\limsup s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le \frac{1}{1-\epsilon}(1+\sqrt{1-\epsilon})^2 \tag{4.41}$$

  almost surely.

If $\alpha_{M_0+1} = 2$, then for sufficiently small $\epsilon > 0$,

$$\alpha_1 > \cdots > \alpha_{M_0} > \alpha_{M_0+1} > 1 + \sqrt{1-\epsilon} > \alpha_{M_0+2} > \cdots > \alpha_M > 1 - \sqrt{1-\epsilon}. \tag{4.42}$$

Hence Theorem 1.1 implies (4.40) but (4.41) becomes

$$\limsup s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le \frac{1}{1-\epsilon}\left(\alpha_{M_0+1} + \frac{(1-\epsilon)\alpha_{M_0+1}}{\alpha_{M_0+1} - 1}\right) \tag{4.43}$$

almost surely.

Therefore (4.37) and (4.40) yield (1.22), and (4.38), (4.41) and (4.43) yield (1.23).

References

[1] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9:611–677, 1999.

[2] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[3] Z. D. Bai and J. W. Silverstein. Exact separation of eigenvelues of large dimensional sample covariance matrices. *Ann. Probab.*, 27(3):1536–1555, 1999.

[4] J. Baik. Painlevé formulas of the limiting distributions for non-null complex sample covariance matrices. in preparation.

[5] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. `http://xxx.lanl.gov/abs/math.PR/0403022`.

[6] A. Buja, T. Hastie, and R. Tibshirani. Penalized discriminant analysis. *Ann. Statist.*, 23:73–102, 1995.

[7] P.J. Forrester. The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402:709–728, 1993.

[8] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[9] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge Press, New York, 1985.

[10] D. Hoyle and M. Rattray. Limiting form of the sample covariance eigenspectrum in PCA and kernel PCA. to appear in proceedings of Neural Information Processing Systems 2003.

[11] K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153:259–296, 2001.

[12] I. Johnstone. On the distribution of the largest Principal Component. *Ann. Statist.*, 29:295–327, 2001.

[13] L. Laloux, P. Cizeau, M. Potters, and J. Bouchaud. Random matrix theory and financial correlations. *Intern. J. Theor. Appl. Finanace*, 3(3):391–397, 2000.

[14] Y. Malevergne and D. Sornette. Collective origin of the coexistance of apparent RMT noise and factors in large sample correlation matrices. `http://xxx.lanl.gov/abs/cond-mat/0210115`.

[15] V.A. Marcenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1:457–486, 1967.

[16] D. Paul. Asymptotic behaviour of the largest eigenvalues of sample covariance matrix when true covariance is a finite perturbation of identity. in preparation.

[17] S. Péché. Universality of local eigenvalue statistics for random sample covariance matrices. *Ph.D. Thesis*, Ecole Polytechnique Fédérale de Lausanne, 2003.

[18] V. Plerous, P. Gopikrishnan, B. Rosenow, L. Amaral, T. Guhr, and H. Stanley. Random matrix approach to cross correlations in financial data. *Phys. Rev. E*, 65(6):066126, 2002.

[19] R. Sear and J. Cuesta. Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.*, 91(24):245701, 2003.

[20] J. W. Silverstein. The smallest eigenvalue of a large dimensional wishart matrix. *Annals of Probability*, 13(4):1364–1368, 1985.

[21] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.

[22] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.

[23] J. W. Silverstein and S. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.

[24] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys*, 108(5-6), 2001.

[25] E. Telatar. Capacity of milti-antenna Gaussian channels. *European transactions on Telecommunications*, 10(6):585–595, 1999.

[26] C. Tracy and H. Widom. Level spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159:33–72, 1994.

[27] C. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177:727–754, 1996.